Online Supplemental Data

Supplemental Methods

Additional details about materials and methods are described here.

Supplemental Tables

Supplemental Table 1: Patient characteristics and sample summary

Supplemental Table 2: Flow cytometry antibodies for sorting of HRS cells

Supplemental Table 3: Significantly mutated genes

Supplemental Table 4: Comparison of the different cHL genetic studies

Supplemental Table 5: Mutational signature analyses

Supplemental Table 6: Significant SCNAs

Supplemental Table 7: Chromosomal rearrangements

Supplemental Table 8: Gene-by-sample matrix

Supplemental Table 9: CCF matrix for all driver events

Supplemental Figures

Supplemental Figure 1: Composition of the cHL dataset.

Supplemental Figure 2: Flow cytometric sorting strategy to isolate HRS cells.

Supplemental Figure 3: CLUMPS analysis and mutation diagrams (lollipop figures) for *PTPN1* and the significantly mutated genes in the cHL cell lines.

Supplemental Figure 4: Supporting data for mutational signature analysis in primary cHLs and cHL cell lines.

Supplemental Figure 5: SCNAs and chromosomal rearrangements in cHL.

Supplemental Figure 6: Comparison of the number of driver events in cHLs with or without *ARID1A* mutations, excluding the two hypermutated cases.

Supplemental Figure 7: MHC class II cell surface expression on HRS cells from tumors with known *MHCII* status.

Supplemental Figure 8: Comparison of mutational density (A) and number of driver events (B) in cHLs with normal vs increased ploidy.

Supplemental Figure 9: Comparison of the mutational density (burden) in EBV^+ and EBV^- cHLs.

Supplemental methods

EBV-encoded small RNA *in situ* hybridization

EBV-encoded small RNA (EBER) *in situ* hybridization was performed as described[1] to determine the EBV status of the primary cHLs where slides were available. All results were reviewed by hematopathologists D.W and J.R.F.

Flow cytometry cell sorting

The HRS cell population was identified as described: 1) intermediate to bright expression of CD30, bright CD40 and CD95; 2) intermediate to bright CD15; and 3) lack of CD20, CD64 and CD5[2,3] (supplemental Figures 2A-B). Patient-matched normal B cells had bright CD19 expression and lacked HRS cell surface markers including CD15 and CD30 (supplemental Figure 2C). All flow cytometric sorting was performed using DIVA 8.0.1 for acquisition and analysis was carried out using FlowJo 10 software.

Library preparation and whole exome sequencing

DNAs were fragmented by Covaris ultrasonication to 250 bp and further purified using Agencourt AMPure XP beads. All samples (HRS and normal B cells) underwent low-input library construction using a Center for Cancer Genome Discovery (CCGD)-modified Kapa HTP kit (quarter reactions of the standard HTP protocol). Libraries were quantified using MiSeq Nano and all samples with successful library preparation (yielding $\underline{\geq}$ 250 ng of DNA libraries) were taken forward for hybrid capture and subsequent sequencing. Normalized libraries were pooled in batches and capture was performed using the Agilent SureSelect Hybrid Capture kit (cat no. G9611A). The captures were further pooled and sequenced over the equivalent of 18 HiSeq 3000 and 3 HiSeq 2500 lanes, respectively, as previously described[4,5]. Following CCGD guidelines of sequence quality metrics (lane performance, relative representation of each sample within a pool and sequencing quality metrics), only samples with ≥80% of the targets having a minimum coverage of 30x were used for further analysis.

Determination of EBV status

Samtools idxstats was used to determine the total number of reads that mapped to the EBV contig (EBV genome build: NC_007605). This number was normalized using the

total number of mapped reads for each cHL case. The normalized EBV count was compared across all cHLs for which the EBV status was known (via EBER) to set the threshold of EBV positive/negative (positive above 0.026, negative below 0.026).

Copy number analysis and calling of significant SCNAs

Copy number analysis was performed by running *ReCapSeq* and *Allelic Capseg* to identify allele-specific copy ratios across the exome. Recurrent SCNAs were identified using *GISTIC2* as previously described[4]. *GISTIC2* determined recurrent SCNAs, however identification of samples with SCNAs in significant focal peaks or chromosome arms was done as post process. For arm-level events, ≥ 50% of the arm had to be covered by segments with sufficient amplitudes (>0.1) to support the alteration. For focal events, ≥ 20% of the region had to be covered by segments that supported (threshold 0.1) the alteration after subtracting all segments in the region by the median segment mean for the arm.

Mutational signature analysis

A. De novo signature discovery in the 23 primary cHLs.

A de novo signature extraction applied to SNVs stratified by 96 tri-nucleotide mutation contexts identified four major mutational processes. The similarity of these signatures to the 30 COSMIC signatures (http://cancer.sanger.ac.uk/cosmic/signatures) was computed with a cosine similarity. The first signature most resembling COSMIC1 (cosine similarity 0.85) was characterized by a superposition of C>T mutations at CpG sites with a background broad spectrum of base substitutions, and C>T/G at T<u>C</u>W (W=A/T) corresponding to the characteristics of APOBEC mutagenesis, suggesting that this signature is an admixture of COSMIC1 and APOBEC signatures. The activity of this signature was pervasive across samples, explaining about 30% overall mutations. The second signature, which did not match any of the known 30 COSMIC signatures with a cosine similarity > 0.75, had characteristic peaks of C>T/G mutations at G<u>C</u>T context corresponding to one of canonical AID known hotspot motifs at RCY (R = A/G, Y= C/T), and preferences for T>A/C/G at TW (W=A/T) context corresponding to non-canonical AID hotspot motifs, and its activity was significantly higher in clustered mutations (62% in clustered mutations) consistent with known AID biology. The third signature most

resembled COSMIC6 (cosine similarity 0.98), which is known to be associated with defective DNA mismatch repairs and found in microsatellite unstable tumors, and its activity was exclusive to a single hyper-mutant tumor (c_cHL_24). The fourth signature was most similar to COSMIC15 (cosine similarity 0.74), another known MSI signature, and its attribution was exclusive to the second hypermutated sample (c_cHL_4).

B. Semi-supervised signature discovery in 21 primary cHLs.

To minimize a possible interference of two MSI signatures with other signatures and enable a separation of the APOBEC signal from COSMIC1, we excluded the two putative MSI samples (c_cHL_24 and c_cHL_4) and repeated a de novo signature extraction for the remaining 21 cHLs, while enforcing two APOBEC signatures (COSMIC2 and COSMIC13) in the signature extraction. For this, we created two artificial samples with a predominant activity of COSMIC2 and COSMIC13 with 10,000 mutations each and added them to the mutation count matrix of 21 primary cHLs. The mutation counts along 96 contexts in the two artificial samples were proportionally distributed according to the normalized profiles of COSMIC2 and COSMIC13 signatures. With the included artificial samples, Bayesian NMF identified two major signatures, COSMIC1 (cosine similarity 0.92) and the AID signature, in addition to the enforced COSMIC2 and COSMIC13 (Figure 2A). Note that the cosine similarity of COSMIC1 was much improved from 0.85 to 0.92 with a separation of APOBEC components. Overall 8% and 5% mutations were attributed to COSMIC2 and COSMIC13, respectively. We used the semi-supervised signature analysis in all downstream evaluations.

C. Semi-supervised signature discovery for the combined cohort of 23 primary cHL and six cHL cell lines.

As described in **B**, we performed a semi-supervised signature discovery for the combined cohort of 23 cHLs and six cHL cell lines (L-1236, L-428, L-540, KM-H2, HDLM-2, SUP-HD1). In addition to the enforced APOBEC signatures (COSMIC2 and COSMIC13), we identified an additional seven signatures (supplemental Figures 4C-D); COSMIC1, AID, and two MSI signatures (COSMIC6 and COSMIC15), and COSMIC 25 (cosine similarity 0.89) and COSMIC11 (cosine similarity 0.94). The attribution of COSMIC11, known to be associated with the treatment with alkylating agents, was exclusive to a single cell line

(SUP-HD1), while the attribution of COSMIC25, which has been identified in cHL cell lines with unknown etiology, was mostly present in three cHL cell lines (KM-H2, L-540, and L-1236) (supplemental Figures 4D-E).

D. Gene-level Signature enrichment analysis.

We annotated each mutation with the probability (likelihood of association) that it was generated by each of the discovered mutational signatures, P_{ms}, where '*m*' denoted a mutation and '*s*' refers to the signature. More specifically, the likelihood of association to the *k*-th signature for a set of mutations corresponding to *i*-th mutation context and *j*-th clustered or non-clustered mutation group was defined as $[\boldsymbol{w}_k\boldsymbol{h}_k / \sum(\boldsymbol{w}_k\boldsymbol{h}_k)]_{ij}$, where $\boldsymbol{w}_k$ and $\boldsymbol{h}_k$ correspond to the *k*-th column vector and *k*-th row vector of $\boldsymbol{W}$ and $\boldsymbol{H}$, respectively. The relative activity enrichment for candidate driver genes in Figure 2D was determined by taking an average of P_{ms} for all mutations in each driver gene.

Supplemental Tables (All provided as separate excel files).

Supplemental Table 1: Patient characteristics and sample summary

Patient characteristics and summary information for each sample is listed including sample ID, morphological subtype, age, gender, EBV status, quality control metrics such as mean target coverage, mean allele fraction and number of coding mutations, purity, ploidy and genetic features.

Supplemental Table 2: Flow cytometry antibodies for sorting of HRS cells

(A) Antibodies are listed with clone, laser channel and manufacturer that were used for the sorting of HRS and normal B-cells from the cHL excision biopsies. **(B)** An expanded antibody panel was used to characterize MHC class II expression on flow-sorted HRS cells. Antibodies are listed with clone, laser channel and manufacturer.

Supplemental Table 3: Significantly mutated genes

(A) Mutated genes ranked by their significance values obtained from *MutSig2CV*. **(B)** Genes with significant spatial clustering within a protein structure as detected by *CLUMPS*.

Supplemental Table 4: Comparison analyses of the cHL genetic alterations identified in this study and those of Tiacci et al[6], Spina et al[7] and Reichel et al[8].

Supplemental Table 5: Mutational signature analyses

(A) Mutational signature activity in the 23 cHLs (de novo analysis), **(B)** Mutational signature activity in the 21 cHLs excluding the two MSI cases.

Supplemental Table 6: Significant SCNAs

(A) List of focal and arm-level SCNAs (copy gain and copy loss) identified with *GISTIC2.0* in the 23 cHLs. **(B)** Summary of genes identified in focal gains. **(C)** Summary of genes in focal losses. **(D)** Summary of COSMIC-defined genes observed in focal SCNAs.

Supplemental Table 7: Chromosomal rearrangements

Summary of chromosomal rearrangements, identified in the 23 cHLs, including chromosomal position of the first and second gene, type of rearrangement, support by the detection algorithm, split read and read pair count as well for the reference and alternate alleles and cancer cell fractions.

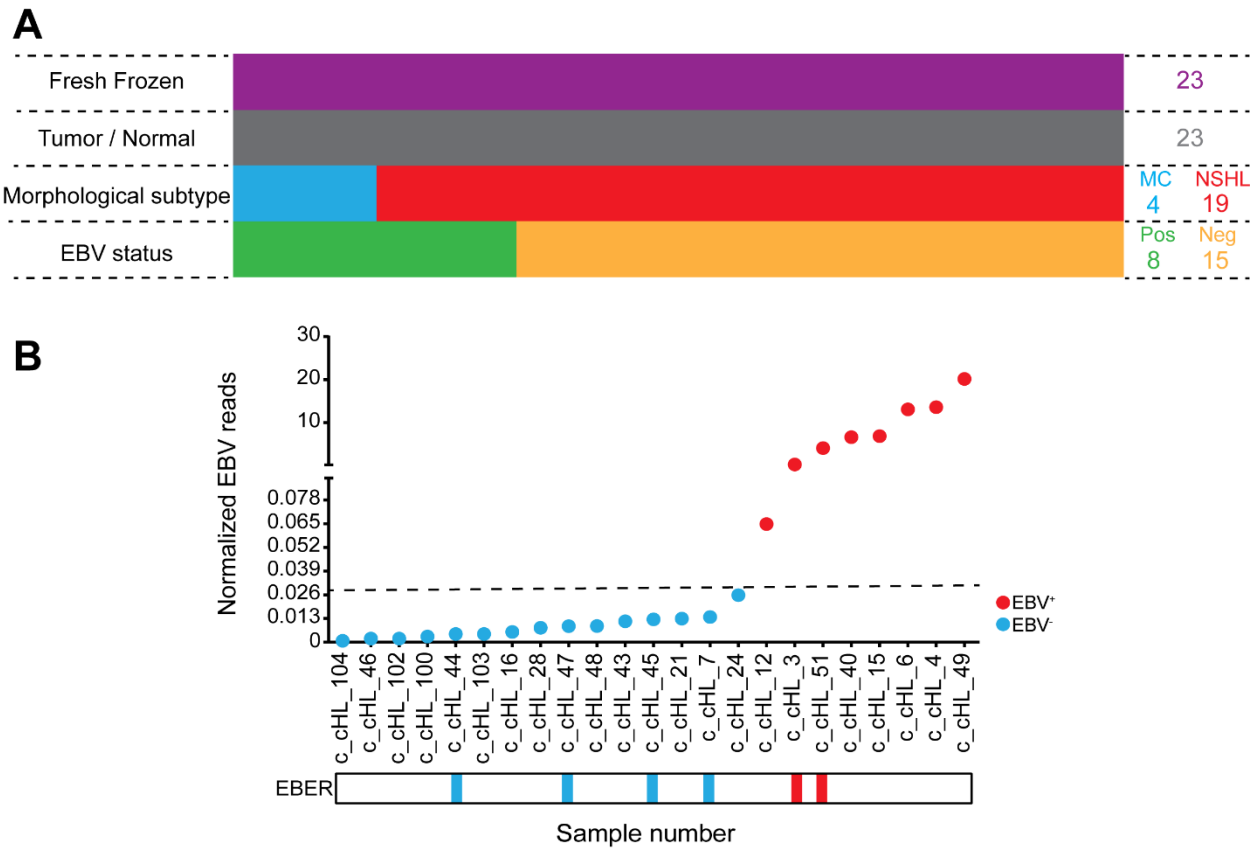
Supplemental Table 8: Gene-by-sample matrix

(A) Gene-by-sample matrix for the 23 cHLs with driver events occurring >2 times are listed. **(B)** Gene-by-sample matrix for 6 cHL cell lines. Mutations (0, absent; 1, synonymous; 2, non-synonymous); SCNAs (0, no SCNA; 1, low level SCNA; 2, high level SCNA); Chromosomal Rearrangements (0, absent; 3, present).

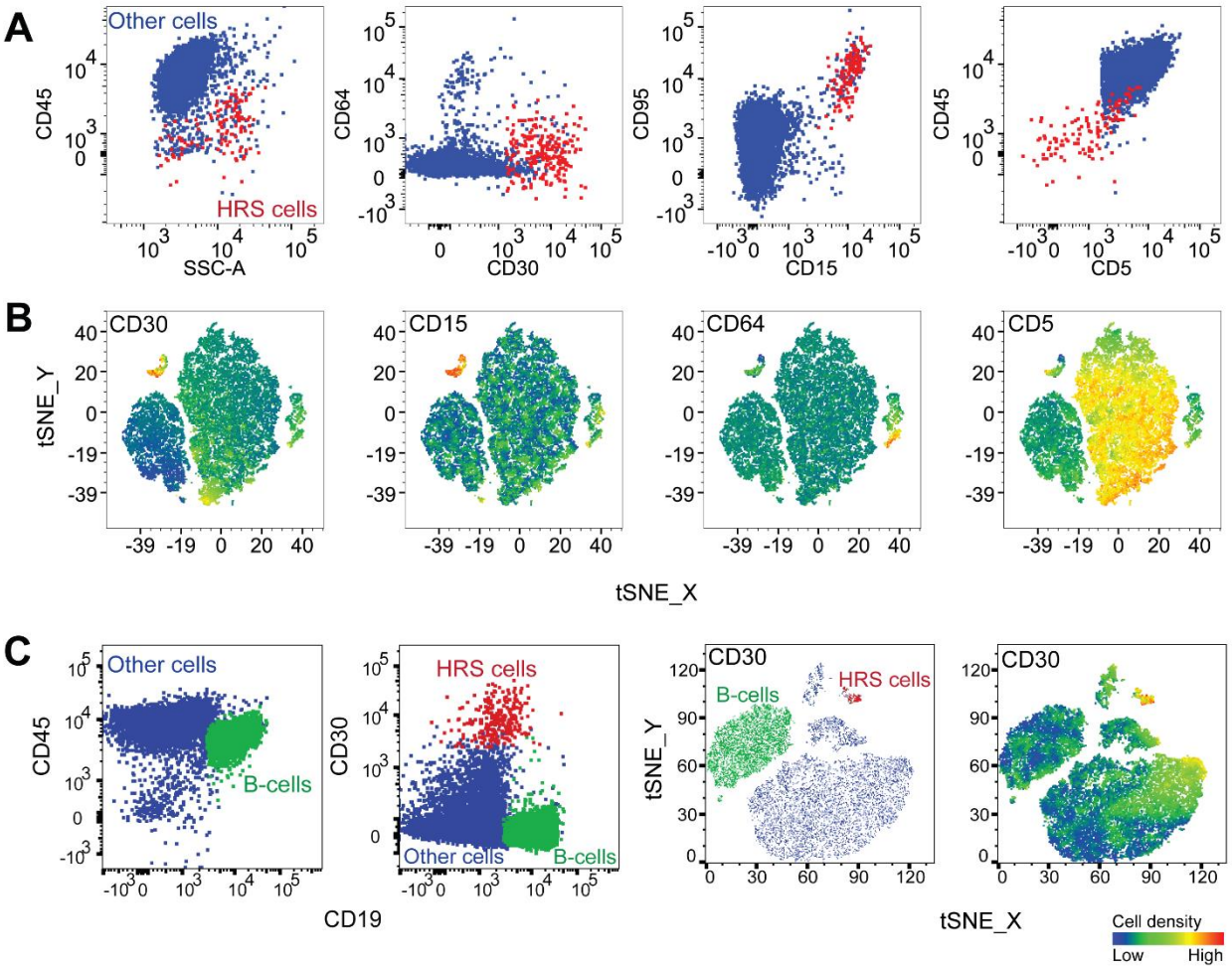
Supplemental Table 9: CCF-matrix for all driver events

CCF matrix for all driver events identified in the 23 cHLs.

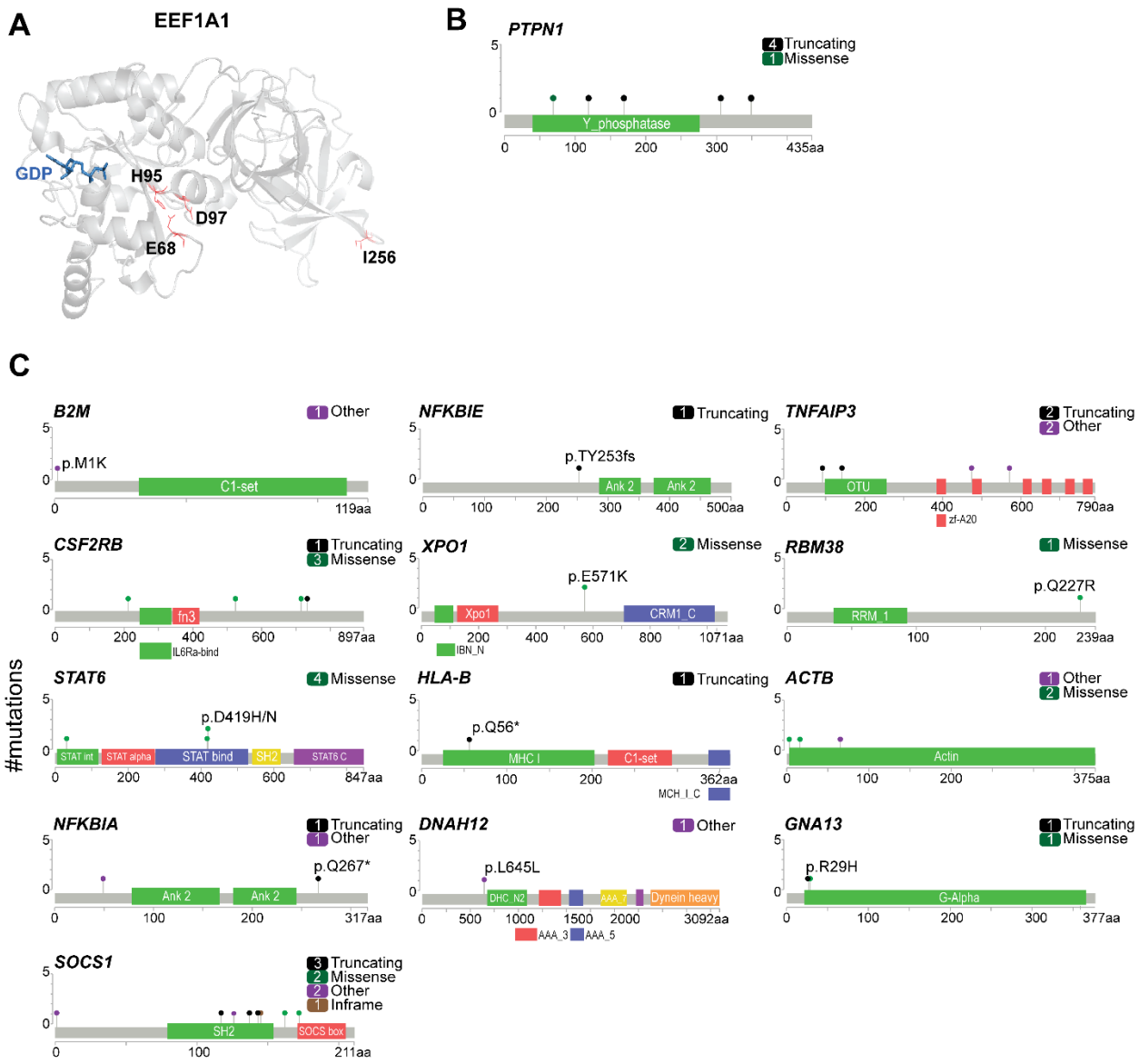
Supplemental Figures



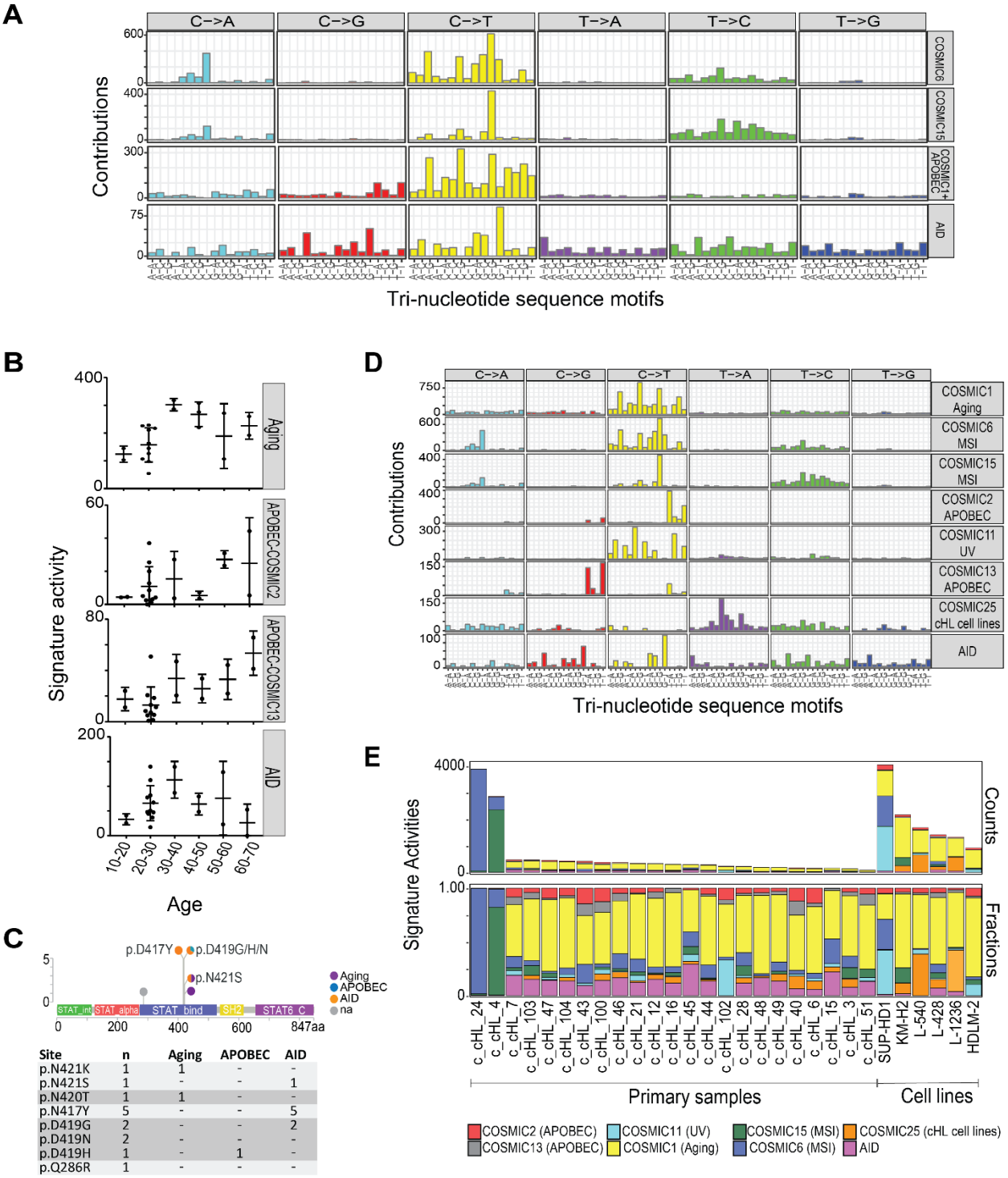
Supplemental Figure 1: Composition of the cHL dataset. (A) The dataset includes fresh frozen excision biopsies with paired normal B-cells from 23 newly diagnosed cHL patients (row one and two). Within the patient cohort, four cHLs were classified as mixed cellularity (MC) and 19 as nodular sclerosis Hodgkin lymphoma (NSHL) (row three) with eight of the 23 patients being EBV^+ and 15 EBV^- (row four). **(B)** The EBV status was determined by the number of reads mapping to the EBV genome (normalized by total mapped reads) and positive/negative EBV status was confirmed in samples with known EBER status. EBV^+ samples are indicated as red dots while EBV^- samples are noted as blue dots.



Supplemental Figure 2: Flow cytometric sorting strategy to isolate HRS cells. (A) HRS cells (indicated in red) are identified by decreased CD45 with intermediate to bright expression of CD30, CD15, CD95, lack of CD64 and CD5. Using the different antibody combinations and sequential gating, other cells (indicated in blue) such as $CD64^+$, $CD5^+$ were excluded to flow cytometric sort a pure HRS cell population. **(B)** The T-distributed Stochastic Neighbor Embedding (tSNE) algorithm was used to visualize the flow cytometry data in a dimension-reduced space and indicate that HRS cells ($CD30^+$, $CD15^+$) are a cell population distinct from the $CD64^+$ and $CD5^+$ cells. **(C)** B-cells (indicated in green) were identified with increased CD19 and lack of CD30 expression. The tSNE plots further confirmed that HRS and B-cells were distinct populations, respectively.

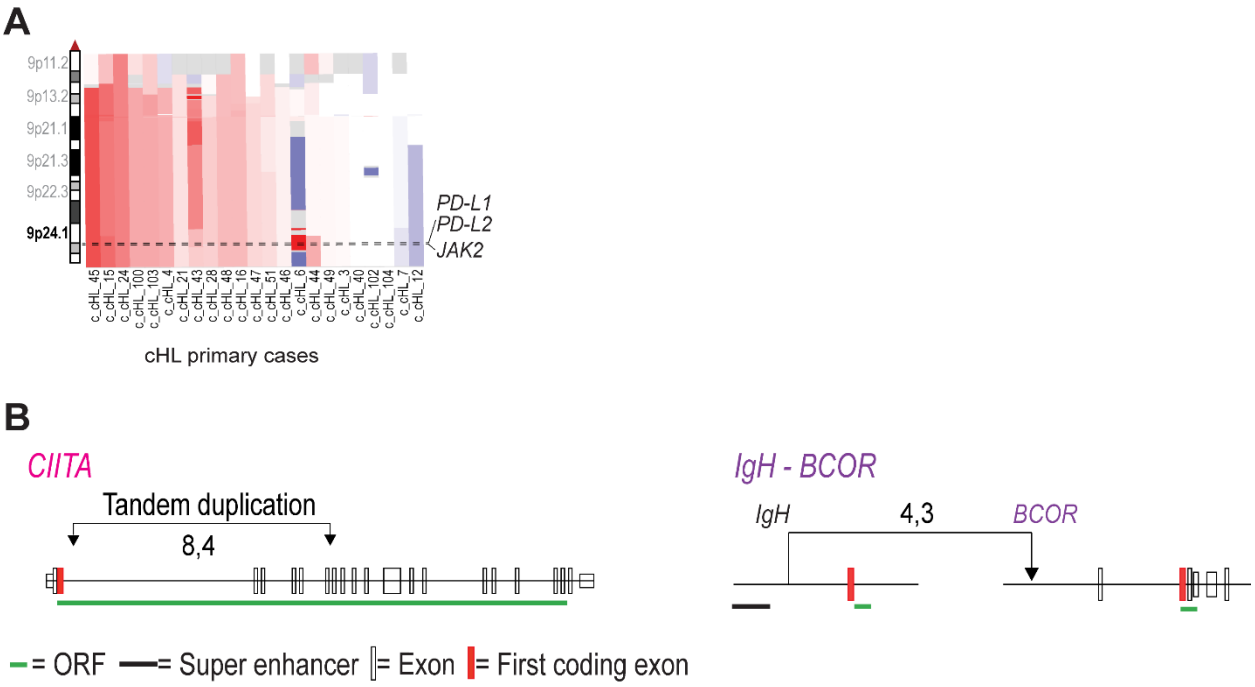


Supplemental Figure 3: *CLUMPS* analysis and mutation diagrams (lollipop figures) for *PTPN1* and the significantly mutated genes in the cHL cell lines. (A) Spatial clustering of mutations identified with *CLUMPS* in *EEF1A1* (PDB: 4c0s). Mutated residues are shown in red and color-intensity and thickness of line scales with number of mutations. Co-crystalized proteins are shown in blue (GDP). **(B)** Mutation diagram for *PTPN1*. **(C)** Significantly mutated genes identified in ≥ 2 cHLs are visualized in the six cHL cell lines (KM-H2, L-428, L-1236, L-540, HDLM-2, SUP-HD1). Positions within the genes are labeled.

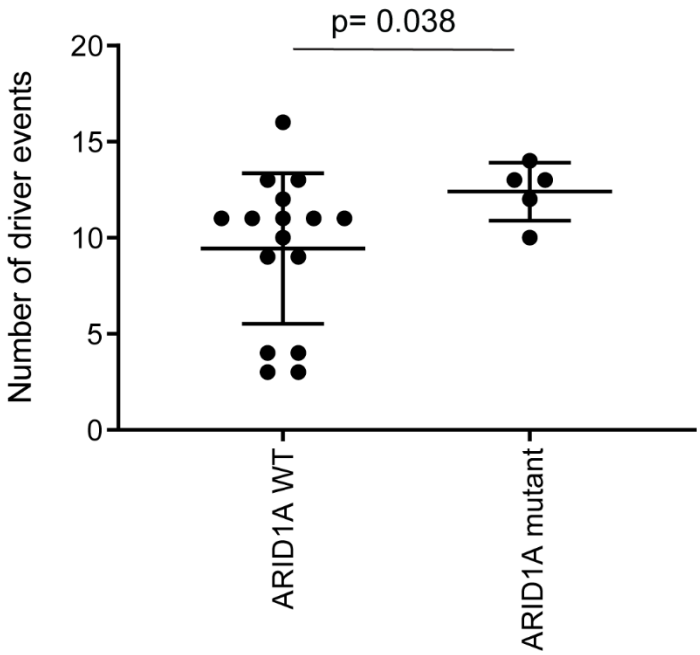


Supplemental Figure 4: Supporting data for mutational signature analysis in primary cHLs and cHL cell lines. (A) De novo signature extraction for the 23 cHLs identified a putative microsatellite instability (MSI) signature (COSMIC6 and COSMIC15) in addition to the two signatures, COSMIC1 with APOBEC and AID. **(B)** Correlations of signature activity to the age at diagnosis across six age groups (n=21 cHLs, excluding the two hypermutated samples). The mean with standard deviation shows the distribution of each signature activity of the samples

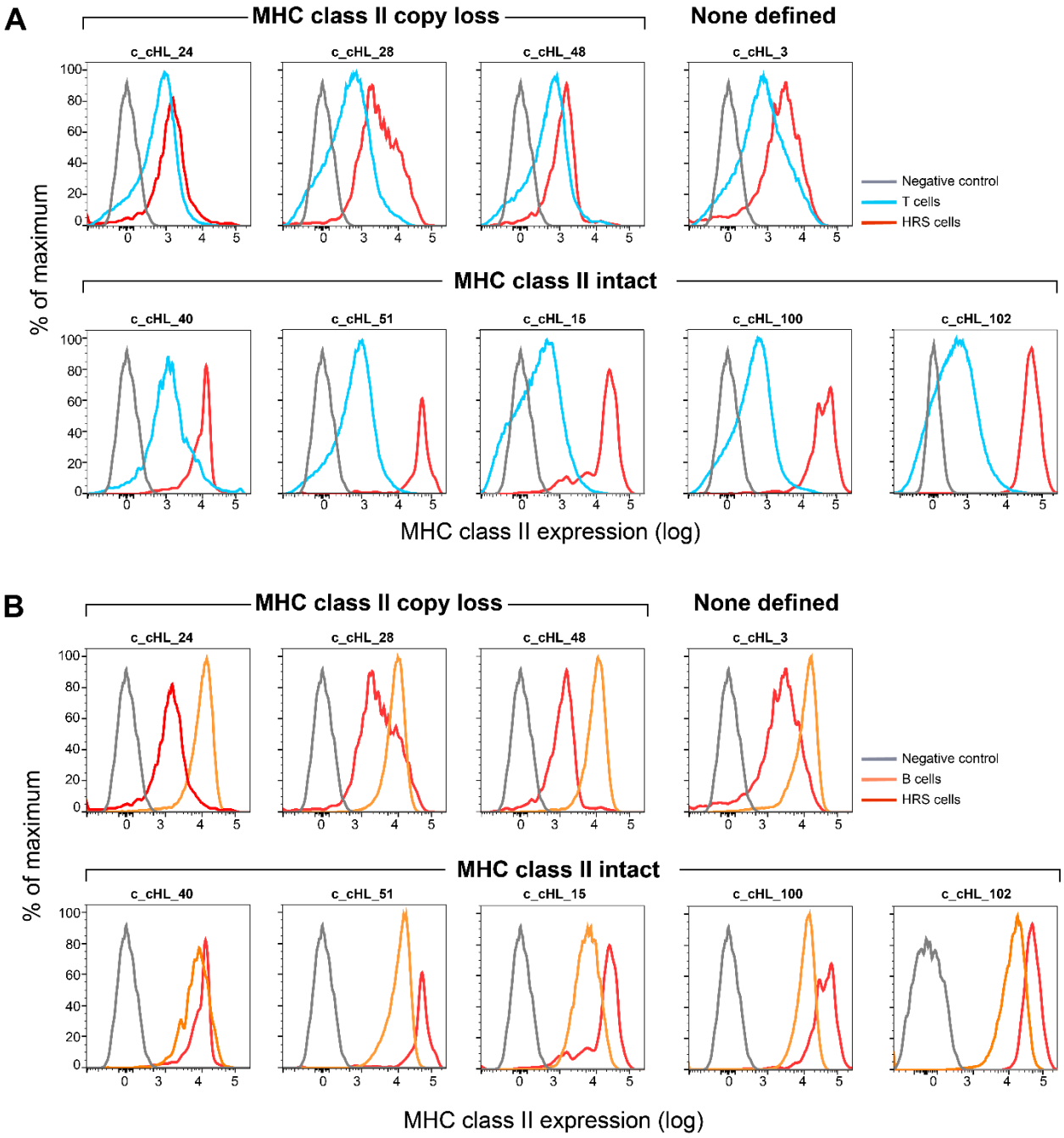
within each age group. Number of cases per age group: 10-20 (n=2), 20-30 (n=11), 30-40 (n=2), 40-50 (n=2), 50-60 (n=2) and 60-70 (n=2). **(C)** Mutational diagram (lollipop figure) for *STAT6* non-synonymous mutations within the functional domains of the protein, visualized using *MutationMapper* v2.1.0[9,10]. The color key denotes the mutational mechanisms (causative probability of the indicated mechanism >0.75) for each site. Below is the table of the STAT6 coding changes and putative underlying mutational mechanisms. **(D)** Mutational signature analysis of the combination of the 23 cHLs and six cHL cell lines. **(E)** Comparison of the different signature activities in the 23 cHLs and six cHL cell lines. The two primary samples with MSI (COSMIC6 and COSMIC15), and COSMIC1 ("Aging"), APOBEC (COSMIC13 and COSMIC2), AID in all samples. COMSIC25 and COSMIC11 (UV) were predominantly seen in the cHL cell lines.



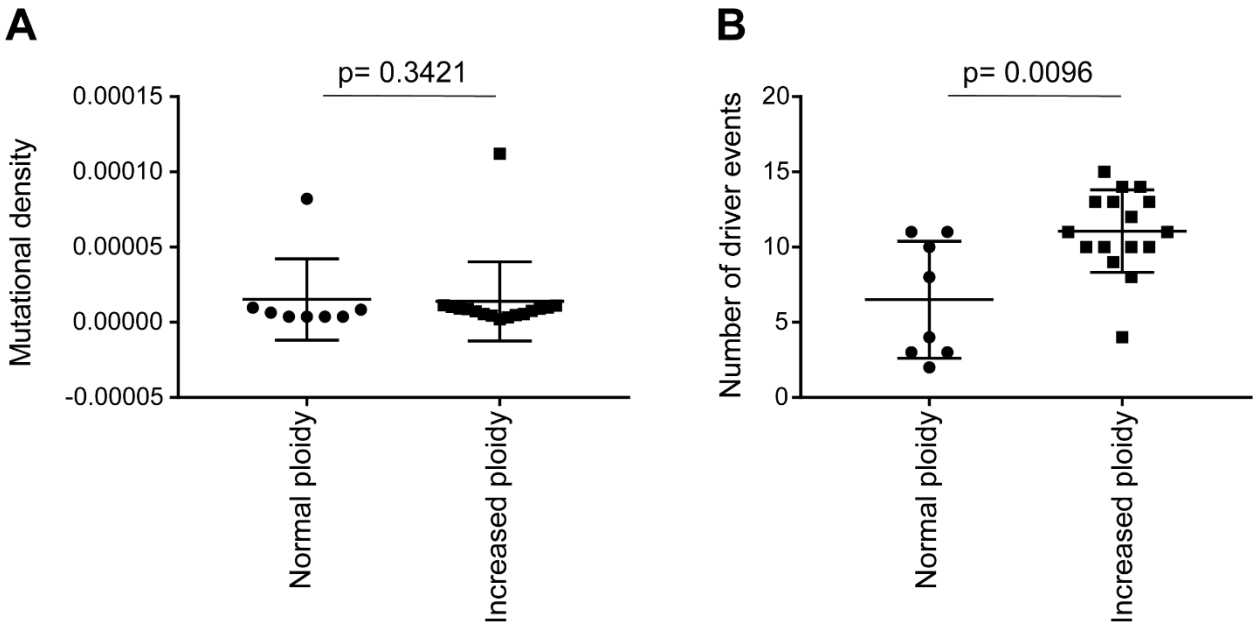
Supplemental Figure 5: SCNAs and chromosomal rearrangements in cHL. (A) IGV plot of copy number from chromosome 9p indicates frequent arm level gain and focal amplification of 9p24.1 (*PD-L1, PD-L2* and *JAK2*). **(B)** SVs in cHL including tandem duplications in *CIITA* and translocation involving *IgH* and *BCOR,* are plotted. Boxes indicate exons including the first coding exon (red); green bar connotes protein-coding exons and intervening introns; and black bar denotes the super enhancer of *IgH*[11]**.**



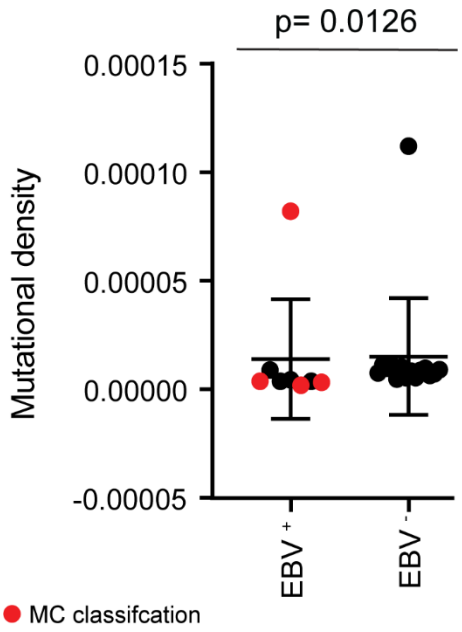
Supplemental Figure 6: Comparison of the number of driver events in cHLs with or without *ARID1A* mutations, excluding the two hypermutated cases. The p-value was obtained using a Mann-Whitney U test.



Supplemental Figure 7: MHC class II cell surface expression on HRS cells from tumors with known *MHCII* status. Flow cytometric analysis of MHC class II expression on purified HRS cells from 9 of the genomically characterized cHLs was performed. MHC class II expression levels on the HRS cells, relative to MHC class II-negative normal infiltrating T cells **(A)** and MHC class II-positive normal B cells **(B)**, was compared in tumors with known *MHCII* status (*MHCII* copy loss in HRS cells, 3 cases; no detectable *MHCII* alterations, 5 cases). HRS cells are shown in red, normal infiltrating T cells in blue **(A)**, normal B cells in orange **(B)** and unstained negative controls in gray. CHLs with decreased HRS cell surface expression of MHC class II were significantly more likely to have *MHCII* copy loss than HRS cells with intact MHC class II expression (p=0.048, Fischer's Exact test) (see also Figure 4E).



Supplemental Figure 8: Comparison of mutational density (A) and number of driver events (B) in cHLs with normal vs increased ploidy. For these comparisons, primary cHLs were separated into two groups based on tumor ploidy (≤2 [normal] versus >2.2 [increased]). The p-values in A and B were obtained using a Mann-Whitney U test.



Supplemental Figure 9: Comparison of the mutational density (burden) in EBV$^+$ and EBV$^-$ cHLs. Primary cHL cases classified as MC are colored red. The p-value was obtained using a Mann-Whitney U test.

References from Supplemental Information

1. Roemer MG, Advani RH, Ligon AH, Natkunam Y, Redd RA, Homer H, Connelly CF, Sun HH, Daadi SE, Freeman GJ, Armand P, Chapuy B, de Jong D, Hoppe RT, Neuberg DS, Rodig SJ, Shipp MA. PD-L1 and PD-L2 Genetic Alterations Define Classical Hodgkin Lymphoma and Predict Outcome. *J Clin Oncol*. 2016;34(23):2690-2697.
2. Fromm JR, Wood BL. Strategies for immunophenotyping and purifying classical Hodgkin lymphoma cells from lymph nodes by flow cytometry and flow cytometric cell sorting. *Methods*. 2012;57(3):368-375.
3. Fromm JR, Thomas A, Wood BL. Flow cytometry can diagnose classical hodgkin lymphoma in lymph nodes with high sensitivity and specificity. *Am J Clin Pathol*. 2009;131(3):322-332.
4. Chapuy B, Stewart C, Dunford AJ, Kim J, Kamburov A, Redd RA, Lawrence MS, Roemer MGM, Li AJ, Ziepert M, Staiger AM, Wala JA, Ducar MD, Leshchiner I, Rheinbay E, Taylor-Weiner A, Coughlin CA, Hess JM, Pedamallu CS, Livitz D, Rosebrock D, Rosenberg M, Tracy AA, Horn H, van Hummelen P, Feldman AL, Link BK, Novak AJ, Cerhan JR, Habermann TM, Siebert R, Rosenwald A, Thorner AR, Meyerson ML, Golub TR, Beroukhim R, Wulf GG, Ott G, Rodig SJ, Monti S, Neuberg DS, Loeffler M, Pfreundschuh M, Trumper L, Getz G, Shipp MA. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med*. 2018;24(5):679-690.
5. Chapuy B, Roemer MG, Stewart C, Tan Y, Abo RP, Zhang L, Dunford AJ, Meredith DM, Thorner AR, Jordanova ES, Liu G, Feuerhake F, Ducar MD, Illerhaus G, Gusenleitner D, Linden EA, Sun HH, Homer H, Aono M, Pinkus GS, Ligon AH, Ligon KL, Ferry JA, Freeman GJ, van Hummelen P, Golub TR, Getz G, Rodig SJ, de Jong D, Monti S, Shipp MA. Targetable genetic features of primary testicular and primary central nervous system lymphomas. *Blood*. 2016;127(7):869-881.
6. Tiacci E, Ladewig E, Schiavoni G, Penson A, Fortini E, Pettirossi V, Wang Y, Rosseto A, Venanzi A, Vlasevska S, Pacini R, Piattoni S, Tabarrini A, Pucciarini A, Bigerna B, Santi A, Gianni AM, Viviani S, Cabras A, Ascani S, Crescenzi B, Mecucci C, Pasqualucci L, Rabadan R, Falini B. Pervasive mutations of JAK-STAT pathway genes in classical Hodgkin lymphoma. *Blood*. 2018;131(22):2454-2465.
7. Spina V, Bruscaggin A, Cuccaro A, Martini M, Di Trani M, Forestieri G, Manzoni M, Condoluci A, Arribas A, Terzi-Di-Bergamo L, Locatelli SL, Cupelli E, Ceriani L, Moccia AA, Stathis A, Nassi L, Deambrogi C, Diop F, Guidetti F, Cocomazzi A, Annunziata S, Rufini V, Giordano A, Neri A, Boldorini R, Gerber B, Bertoni F, Ghielmini M, Stussi G, Santoro A, Cavalli F, Zucca E, Larocca LM, Gaidano G, Hohaus S, Carlo-Stella C, Rossi D. Circulating tumor DNA reveals genetics, clonal evolution, and residual disease in classical Hodgkin lymphoma. *Blood*. 2018;131(22):2413-2425.
8. Reichel J, Chadburn A, Rubinstein PG, Giulino-Roth L, Tam W, Liu Y, Gaiolla R, Eng K, Brody J, Inghirami G, Carlo-Stella C, Santoro A, Rahal D, Totonchy J, Elemento O, Cesarman E, Roshal M. Flow sorting and exome sequencing reveal the oncogenome of primary Hodgkin and Reed-Sternberg cells. *Blood*. 2015;125(7):1061-1072.
9. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404.
10. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.
11. Chapuy B, McKeown MR, Lin CY, Monti S, Roemer MG, Qi J, Rahl PB, Sun HH, Yeda KT, Doench JG, Reichert E, Kung AL, Rodig SJ, Young RA, Shipp MA, Bradner JE. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell*. 2013;24(6):777-790.